

## DOE Bioenergy Technologies Office (BETO) 2023 Project Peer Review

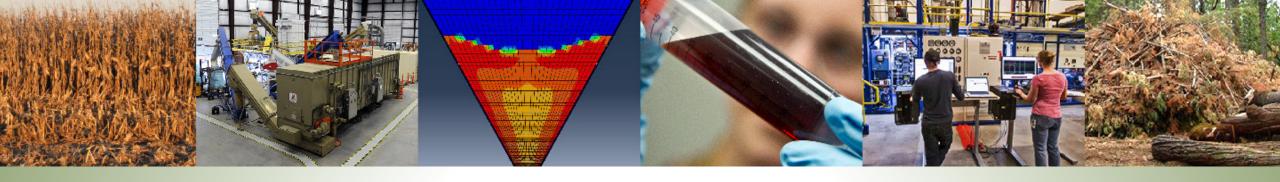
FCIC Task 4 – Data Integration and Web Portal Development

**April 6, 2023 Feedstock-Conversion Interface Consortium (FCIC)** 

Jim Collett – PNNL Rachel Emerson - INL



This presentation does not contain any proprietary, confidential, or otherwise restricted information



## **Project Overview**



## FCIC Task Organization



Feedstock

Preprocessing

Conversion

Task 2: Feedstock Variability

Task 5: Preprocessing

Task 6: High-Temperature Conversion

**Task 1: Materials of Construction** 

Task 7: Low-Temperature Conversion

**Task 3: Materials Handling** 

**Enabling Tasks** 

**Task X: Project Management** 

**Task 4: Data Integration** 

Task 8: TEA/LCA
Task 9: FMEA

**Task X: Project Management:** Provide scientific leadership and organizational project management

**Task 1: Materials of Construction:** Specify materials that do not wear, or break at unacceptable rates

**Task 2: Feedstock Variability:** Quantify & understand the sources of biomass resource and feedstock variability

**Task 3: Materials Handling:** Develop tools that enable continuous, steady, trouble free feed into reactors

**Task 4: Data Integration:** Ensure the data generated in the FCIC are curated and stored – FAIR guidelines

**Task 5: Preprocessing:** Enable well-defined and homogeneous feedstock from variable biomass resources

Task 6 & 7: Conversion (High- & Low-Temp Pathways): Produce intermediates for further processing

**Task 8:Crosscutting Analyses TEA/LCA:** Valuation of intermediate streams & quantify variability impact

Task 9:Failure Mode & Effects Analysis (FMEA): Standardized approach for assessing attribute criticality



## Task 4 Project Overview

feedstock-conversion interface consortium

- Objective: To provide a web-enabled database and public web portal for integrating, standardizing, and distributing FCIC data to industry stakeholders.
- Impact: This Task will accelerate the design and deployment of biorefineries and support their profitable operation by providing a "one-stop shop" for upto-date data and information on understanding and mitigating the impacts of feedstock variability on bioenergy conversion processes.
- Outcome: An online community of industry and DOE stakeholders sharing data, knowledge, and tools via the FCIC Bioenergy Data Hub with a target of having 100+ active users from commercial entities.







#### Industry Stakeholders



Feedstock Producers



Equipment Designers



Biorefinery Integrators



R&D Teams



### The Task 4 Team





Jim Collett **PNNL** Task Lead Site Architecture: **Data Tool** Integration; Task & Subtask Content; Planning and Budgeting; Low Temperature Conversion SME\*



Rachel Emerson INL Task Co-Lead Site Architecture: Case Study Web Design & Publishing; FMEA\*\* Integration; Interface to Bioenergy Feedstock Library; Feedstocks SME



**Bob Kinoshita** INL Data Tool Design; JavaScript, Python, and R Coding: Preprocessing SME



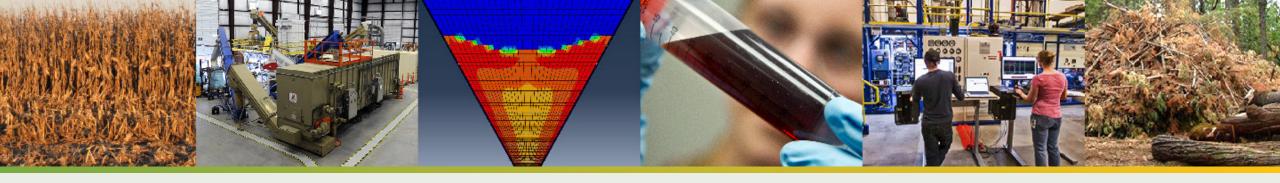
Anne Starace **NREL** Case Study Web Design & Publishing; High **Temperature** Conversion SME



Shaun O'Leary PNNL AWS\*\*\* Cloud Infrastructure; Site Security; DOE **Policy Compliance** 



Matt Macduff **PNNL** LabKey Server and Database Administration

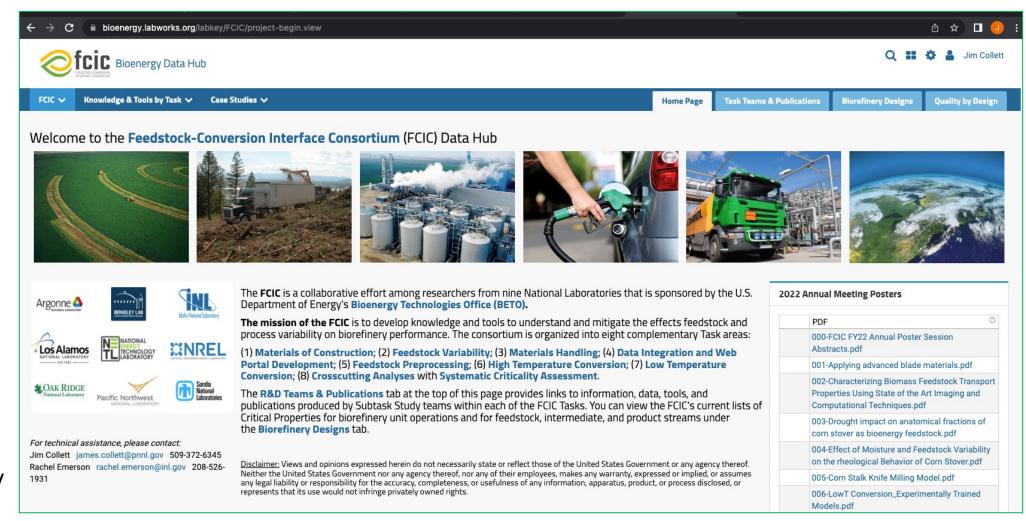






The Data Integration and Web Portal Task targets FY21 Peer Review recommendations to:

- Develop innovative ways to immerse the national laboratories with industry.
- Establish quality specifications on feedstocks
- Aggressively promote FCIC results and visibility



The FCIC Bioenergy Data Hub is now online for invited user testing at <a href="https://bioenergy.labworks.org/labkey/FCIC">https://bioenergy.labworks.org/labkey/FCIC</a>























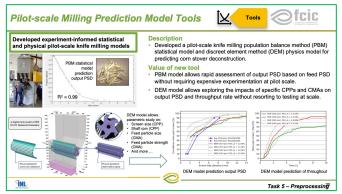
- Task 1. Materials of Construction
- Task 2.
   Feedstock Variability
- Task 3. Materials Handling
- Task 5.
   Feedstock
   Preprocessing
- Task 6.High-Temperature Conversion
- Task 7. Low Temperature Conversion
- Task 8. Cross-Cutting Analysis and Systemic Criticality Assessment
- Task 9. Failure Mode & Effects Analysis (FMEA):



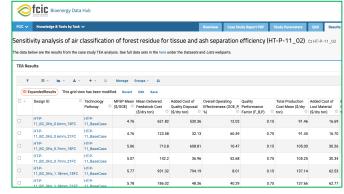
FCIC Case Studies (5)



Journal Articles (78)

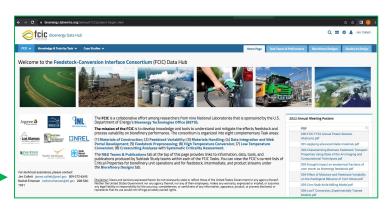


Knowledge and Tool Slides (135)



Supporting Datasets

 Task 4. Data Integration & Web Portal Development



The FCIC Bioenergy Data Hub provides industry stakeholders with integrated access to Case Studies, Journal Articles, Knowledge and Tool Slides and Technical Reports produced by FCIC Researchers from 9 National Labs (NLs) who collaborate within 34 subtask teams within 8 experimental and analytical tasks.





#### Data integration and enhancement across publications to increase their value to stakeholders

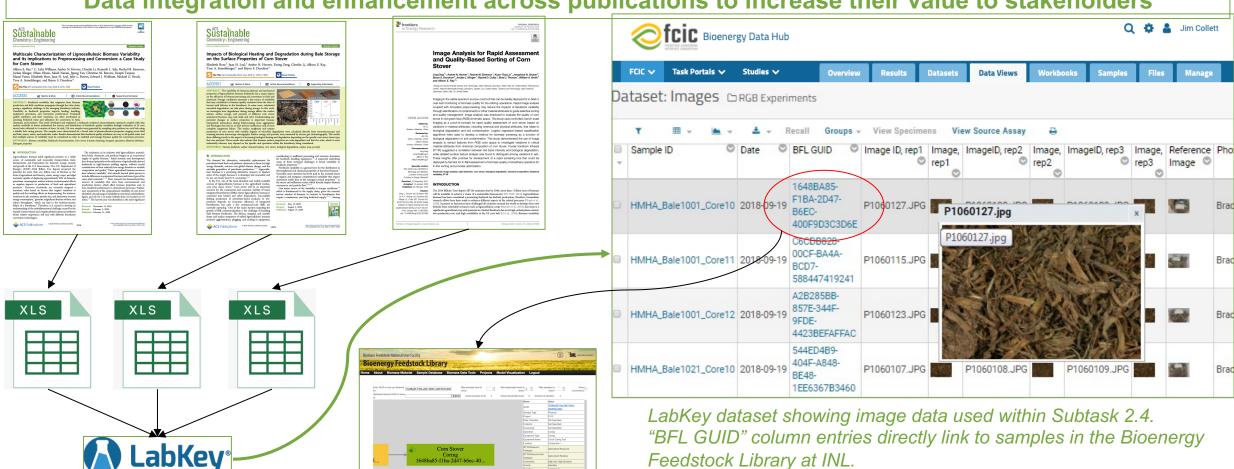


Image Analysis for Rapid Assessment and Quality-Based Sorting of Corn Stover. Ling, et al. FRONTIERS IN ENERGY RESEARCH, 2022.

https://doi.org/10.3389/fenrg.2022.837698



#### **Technical Approach**

Deploy the FCIC Bioenergy Data Hub using the open source <u>LabKey web application platform</u> to make FCIC data and information Findable, Accessible, Interoperable, and Reusable (FAIR) in accordance with the DOE Office of Science and Technology Information's <u>Research Data Management Policy</u>.

#### **Challenges**

- Incentivizing our R&D teams to use the FCIC-branded Data Hub as a "one-stop-shop" for connecting with stakeholders to
  resolve feedstock-conversion interface challenges. We have met this challenge by finding and supporting Data Hub champions
  at the each the member National Labs.
- Harmonizing data terms and units of measure across FCIC Case Studies to create value for stakeholders who now must parse such information from a variety of journals, NL technical reports, and PowerPoint files on the Internet.

#### **Metrics**

- Technical: User Activity Tracking: numbers of Data Hub site visitors once Data Hub is fully public; variety and quantity of FCIC information products on the site; number of log-ons to view/download new FCIC knowledge products when the are promoted via email or other channels.
- Impact: Focus groups and surveys at conferences, on the Data Hub itself, and emailed directly to stakeholders.

#### **Risks/Mitigation Strategies**

- Sustained commitment from FCIC Leadership to support common data standards and to encourage timely data uploading, integration, and dissemination are essential for engendering stakeholder confidence in the FCIC.
- Meeting User Experience expectations and maintaining near 100% uptime are necessary to turn new Data Hub users into devoted fans of FCIC knowledge products and tools.



#### Risks/Mitigation Strategies (continued)

- LabKey Premium Edition Server Software on the FCIC Data Hub provides enterprise-grade, commercial-quality web
  application software deployed on the Amazon Web Services cloud for excellent uptime and availability with same-day technical
  support and weekly meetings with LabKey software engineers.
- Datasets and code are compatible with the free, open-source Community Edition of LabKey Server to allow legacy access to data and information if funding for the consortium is discontinued.
- Separate "production" and "development" servers are maintained by a professional IT team at PNNL to minimize risks of code deployment and software upgrades.

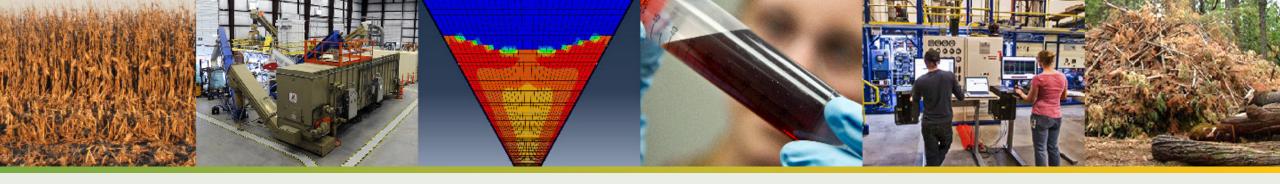
#### Communication/Collaboration

• Two Task 4 Team meetings per month; Collaboration between Task 4 Team and experimental and analytical task teams to standardize data terms and units of measure and to development web content for Case Studies hosted on the Data Hub.

#### DEI

• The LabKey platform has been successfully used to support DEI outreach to high school science classes in other BETO-funded work (2.3.4.106) and may be leveraged within the FCIC to support DEI within Task X.





## 2 - Progress and Outcomes



# 2 – Progress – FCIC Data Stakeholder Advisory Panel provided guidance on Data Hub Design and Content



- Brandon Emme, FCIC IAB, Head of Engineering and Operations, Technology Development at ICM, Inc.
- **Glenn Farris,** FCIC IAB, Executive Vice President Of Operations at Lee Enterprises Consulting, Inc.
- Mel Koch, Principal Scientist, Assoc. Dir., UW Center for Process Analysis and Control; former Global Director of Analytical Services at Dow
- Ray Chrisman, Principal, Chrisman Consulting; Affiliate Faculty, UW Forest Resources; former Technology Manager at Dow
- Mark Penshorn, Lead Chemical Engineer at ICF;
   2021 FCIC Peer Review Panel Lead
- Steve Hanson, Account Manager, Director of User Education, LabKey Software
- Adam Rauch, Vice President of Product Strategy, LabKey Software

Data Hub is now open to non-DOE users on a limited basis for beta testing. Full opening planned for FY23.

#### **Topics Considered**

- FCIC Data Integration and Sharing
  - Data Hub Overall Design
  - Task/Subtask Teams & Workflow Vision
  - Data Finder
  - Quality by Design for Biorefineries
  - Case Studies
  - Building a community of bioeconomy stakeholders

#### **Feedback Summary**

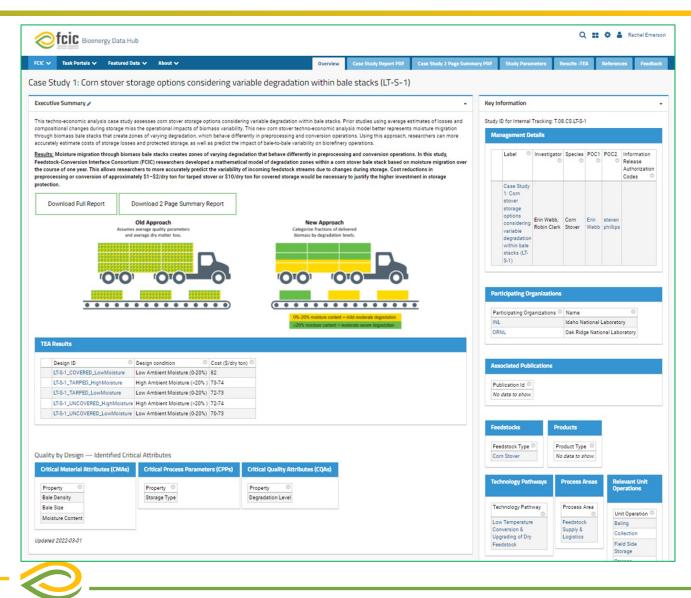
- Users will be interested in specific questions (such as ash content before and after a given unit operation)
- Suggested integration of data beyond FCIC studies
- Make strong connection to BFL and Phyllis2
- Ensure clear navigation, help, and tutorials
- Analytical methods behind data essential
- Data Hub is a great way to get to know the researchers behind FCIC work

# 2 – Progress – New Case Study Content and Redesign for Integrated Data Access



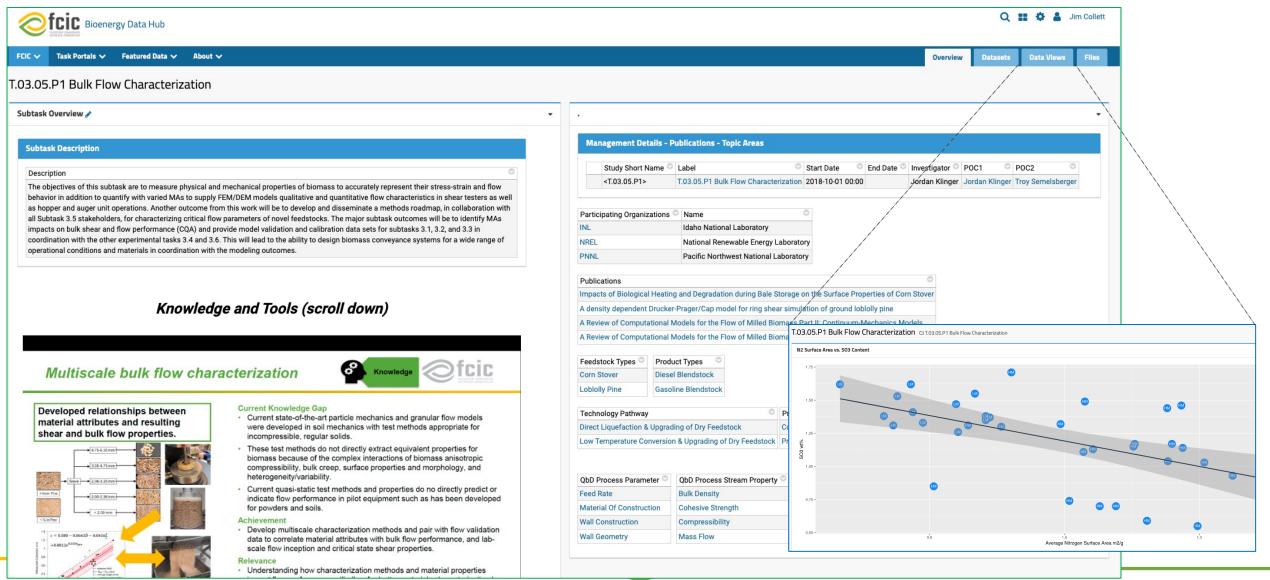
Task 4 worked with Task 8 and FCIC leadership to harmonize Case Study data and redesign them for integrated access on the Data Hub:

- Biomass Deconstruction and Fermentation Performance based on Isolated Anatomical Fractions (LT-C-5)
- Corn stover storage options considering variable degradation within bale stacks (LT-S-1)
- Effect of Tree Age and Fraction on Fast Pyrolysis (HT-C-7)
- Particle scale impacts on deconstruction energy (HT-P-5).
- Sensitivity analysis of air classification of forest residue for tissue and ash separation efficiency (HT-P-11\_02)



# 2 – Progress – Improved design and dynamic content to introduce stakeholders to FCIC Subtask Teams

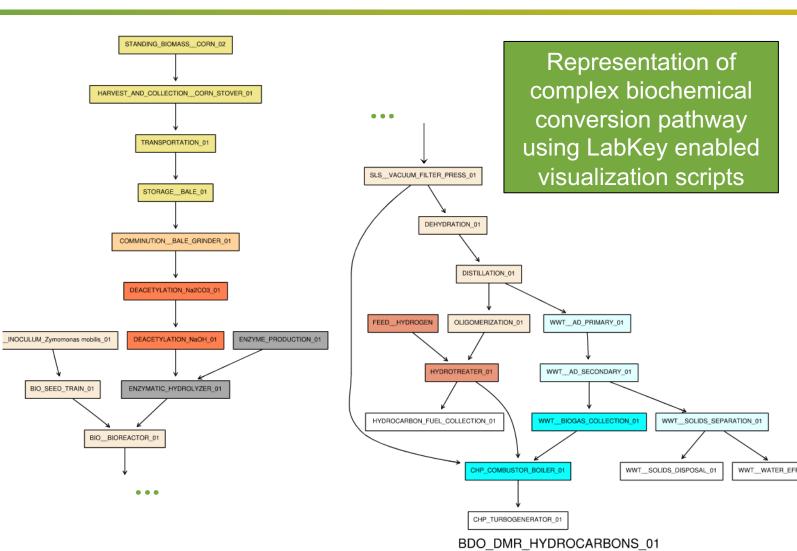




# 2 – Progress – Rapid composition and display of technology pathways from QbD data tables







#### **Description**

A new tool on the <u>FCIC Data Hub</u> enables users to rapidly compose and visualize technology pathways from a standardized set of unit operations linked to FCIC data.

#### Value of new tool

This new tool organizes and presents biorefinery designs, unit operations, and performance data within a standardized Quality by Design (QbD) framework. Industry stakeholders save time by viewing and downloading FCIC data that have been made Findable, Accessible Interoperable, and Reusable (FAIR).

#### **Industry Impact**

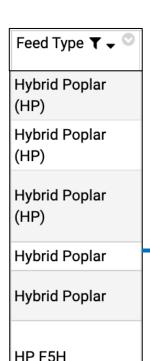
The technology pathway composition and wit\_water\_effuent\_rendering tool has been deployed on the Data Hub. Work is underway to display hyperlinked QbD data on the pathway diagrams.



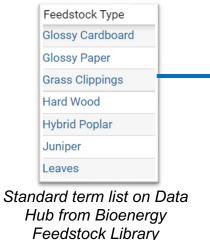
# 2 – Progress – Automated harmonization of data terms using the Record Linkage Toolkit

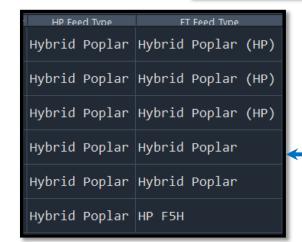




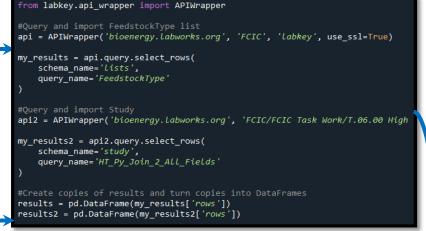


Nonstandard terms in datasets coming from multiple experimental teams





Nonstandard terms are automatically mapped to standard terms



Python script

(Devin Power, PNNL)

Index	Index	Feed Type				
6	8	1				
6	58	1				
6	59	1				
6	60	1				
6	61	1				

Indexing and Classification



The **Python Record Linkage Toolkit\*** is a library to link records in or between data sources. The toolkit provides most of the tools needed for record linkage and deduplication. The package contains indexing methods, functions to compare records and classifiers. With it you can:

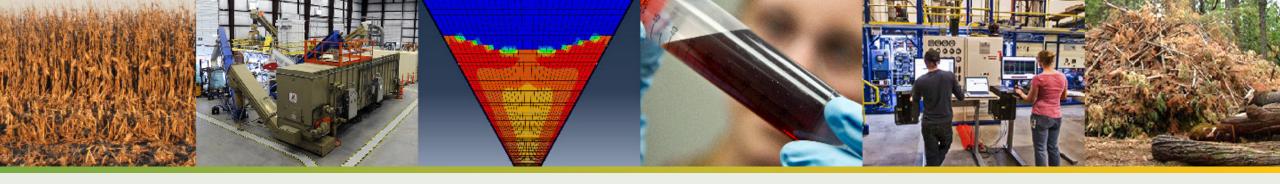
- Clean and standardize data.
- Make pairs of records using smart indexing methods such as blocking and sorted neighborhood indexing
- •Compare records using many comparison and similarity measures for different types of variables such as strings, numbers and dates.
- •Perform classifications using both supervised and unsupervised algorithms.

\*https://recordlinkage.readthedocs.io/en/latest/about.html
© Copyright 2016-2022, Jonathan de Bruin Revision bd5cd08a
Free, open-source code under GPLv3 and BSD-3-Clause licenses









## 3 – Impact



## 3 – Impact



- Industrial stakeholders will have access to:
  - A modern, commercial-quality collaborative platform to integrate, standardize, and archive datasets for advanced analytics and TEA/LCA.

\_

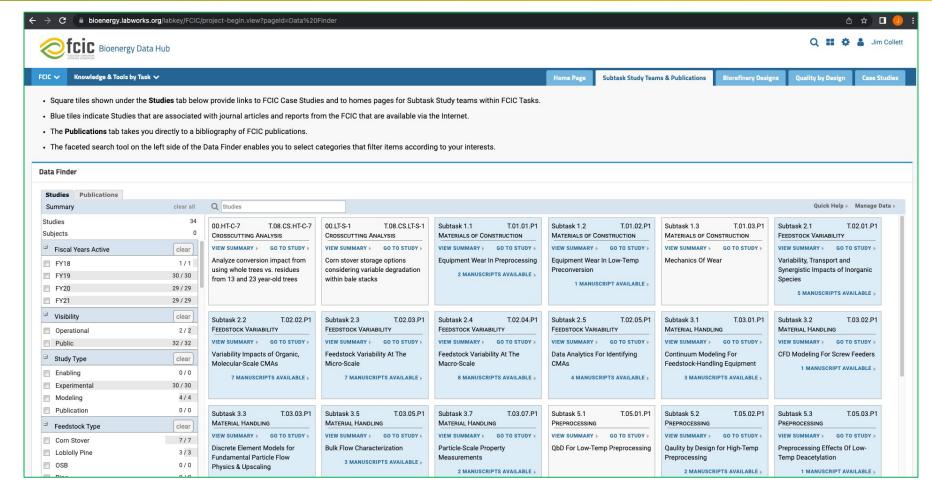
- A web portal with <u>advanced search tools for sharing FCIC datasets</u> with industrial stakeholders.
- A "one-stop-shop for accessing FCIC Case Studies, journal articles, datasets, and info briefs (i.e., K&T slides).
- The Data Hub's QbD database and searchable Subtasks and publications will provide welldocumented data and evidence for:
  - Material properties for biorefinery feedstocks, intermediates, and products.
  - Process parameters for commercial unit operations to facilitate biorefinery scale-up.
  - Data that can support market specifications for feedstocks, intermediates, and products throughout the value chain.



## 3 - Impact - Dissemination



- We will promote the <u>FCIC</u>
   <u>Data Hub</u> as the "go-to"
   platform for datasets on
   feedstock variability and
   biorefinery unit operations via
   webinars, publications, and
   conference presentations.
- High quality, curated datasets on the Data Hub will be convincing "calling cards" for industrial partnerships that lead to commercialization and deployment of BETOfunded research.



Data Finder interface on the home page of the FCIC Data Hub at

https://bioenergy.labworks.org/labkey/FCIC/project-begin.view?pageId=Data%20Finder



## 3 – Impact – Al innovations such as ChatGPT create new opportunities and risks for FCIC data access

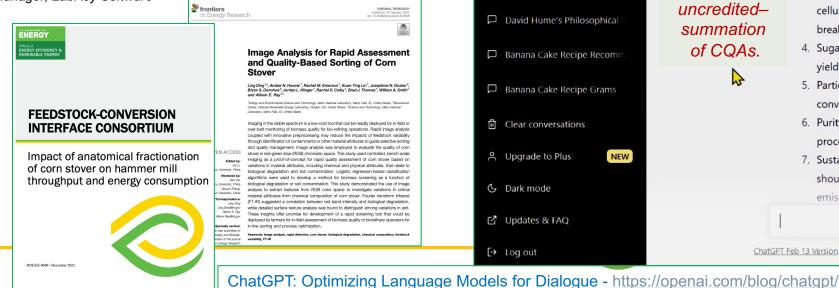




A ChatGPT response to the question "What are the critical quality attributes [CQAs] when converting corn stover into biofuel?"\* suggested that their algorithm drew heavily from FCIC publications that are freely available on the Internet.

We are examining the benefits and risks of such online AI tools to enhance the accessibility and utility of FCIC data and information while maintaining stakeholder confidence in our work.

\*Performed in collaboration with Steve Hanson, Technical Account Manager, LabKey Software





## 3 – Impact – The FCIC is democratizing access to harmonized bioenergy data to accelerate profitable biorefinery deployment



#### FY21 Peer Review Feedback:

"The project is very ambitious in the amount of data and information it is encompassing. Standardizing QbD nomenclature as well as linking to other programs such as the Bioenergy Feedstock Library will be key to institutionalizing the concept across FCIC."

"...the challenge will be to get industry and even other technology areas of BETO to adopt the standardized nomenclature and start using the data sets and reports."



Center for Open Science's strategy for culture change



https://www.cos.io/blog/strategy-for-culture-change

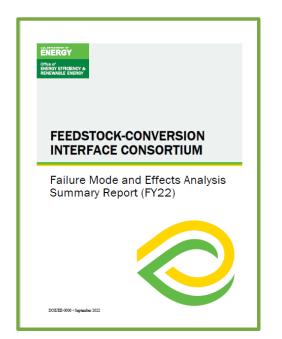
## 3 – Impact – Supporting FCIC Publications and Report



The Data Hub's LabKey infrastructure was used to support Failure Mode Effects Analysis that was performed under Task 9, resulting in the following publication:

 Emerson, Rachel M, Burli, Pralhad Hanumant, Vega Montoto, Lorenzo J, and Bhattacharjee, Tiasha. FEEDSTOCK-CONVERSION INTERFACE CONSORTIUM: Failure Mode and Effects Analysis Summary Report (FY22). United States: N. p., 2022. Web. doi:10.2172/1894327.







## Summary



#### **Technical Approach**

Deploy the FCIC Bioenergy Data Hub using the open source LabKey web application platform to make FCIC data and information Findable, Accessible, Interoperable, and Reusable (FAIR) in accordance with the DOE Office of Science and Technology Information's Research Data Management Policy.

#### **Impact**

The 3-year outcome of this Task will be to establish an online community of industry of stakeholders, National Lab researchers, and BETO staff sharing data, knowledge, and tools via the FCIC Data Hub. All FCIC Integrated Analysis Task Case Studies, publications, QbD properties and supporting information data produced by the FCIC will be available to industry stakeholders via the Data Hub, with a target of having 100+ active users from commercial entities.

#### **Achievements**

- The FCIC Data Hub has been opened to non-DOE users on a limited basis for beta testing. It contains 8 Case Studies with QbDharmonized datasets ready for downloading; one-stop bibliographic access to more than 80 FCIC journal articles and technical reports; and 135 Knowledge & Tool slides. Full opening planned for FY23.
- The FCIC Data Stakeholder Advisory Panel was convened for guidance on Data Hub design, function, and content.
- Improved web design and dynamic content tools created to introduce stakeholders to FCIC Subtask Teams and their work.
- New Tools for rapid composition and dynamic display of biorefinery Technology Pathways from unit operation and QbD data tables.
- Automated Harmonization to Standard Data Terms using the Record Linkage Toolkit
- Analysis of how 3<sup>rd</sup>-party online AI innovations such as ChatGPT will influence interpretation of FCIC knowledge products.

### **Quad Chart Overview**



#### **Timeline**

- October 1, 2021
- September 30, 2024

	FY22 Costed	Total Award				
DOE Funding	\$550K	\$1650K				
Project Cost Share *	NA	NA				

TRL at Project Start: 2 TRL at Project End: 8

### **Project Goal**

Provide a modern, collaborative computational environment for hypothesis development, experimental and modeling workflow management, integration of datasets and metadata, and deliverables sharing between FCIC subtasks within a uniform Quality by Design framework, as well as a portal for public access to FCIC results, data, and software.

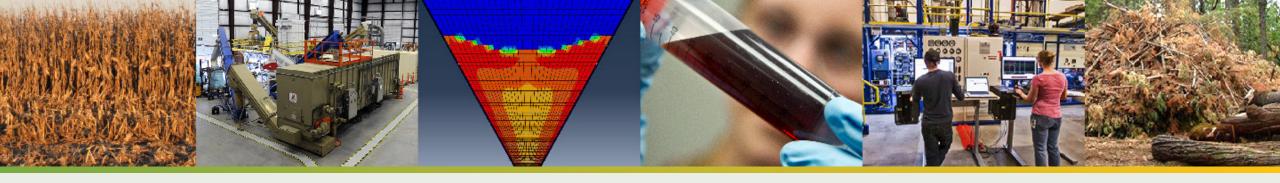
### End of Project Milestone

Realization of an online community of industry stakeholders, National Lab researchers, academic partners, and BETO staff who share data, knowledge, via the FCIC Data Hub with a target of having 100+ active users from commercial entities.

### Funding Mechanism

2021 Lab Call – FCIC Merit Review **Project Partners\*** 





## **Additional Slides**



## 2 – Progress – Task 4 Milestones



FY22 Q1 (PNNL) Data Hub Ingestion and Harmonization of Task 8 Data Sheets; form Data Stakeholder Advisory Team	Harmonization and database ingestion of parameters, data types, and units of measure for Case Study Data Information Sheets generated by Task 8-led planning for FY22 TEA with Experimental Task Teams. Data Stakeholder Advisory Team formed for feedback and guidance on industry relevance of Data Hub resources, formats, and standards.	12/31/21 Complete
FY22 Q2 (INL): Case Studies Data Hub Integration and Publication	Integrate 2 Preprocessing focused Case Studies (batch 1) into Data Hub to support Task 8 presentation of live data tables, visualizations, and narratives to Data Stakeholder Advisory Team.	3/31/22 Complete
FY22 Q3 (NREL): Case Studies Data Hub Integration and Publication	Integrate 2 Conversion focused Case Studies (batch 2) into Data Hub to support Task 8 presentation of live data tables, visualizations, and narratives to Data Stakeholder Advisory Team.	6/30/22 Complete
FY22 Q4 (PNNL): Data Stakeholder Workshop	Inaugural annual Data Stakeholder workshop and training conducted using the Data Hub live for Case Study reviews, platform evaluation and feedback, and training. (delayed while National Lab information release processes are being harmonized)	9/30/22 Delayed
FY23 Q1 (INL) Integration of FY22 material Needs with core FCIC datasets and metadata	Integration of FY22 Material Needs with core FCIC datasets and metadata for preprocessed feedstocks within the Data Hub to link data to R&D tasks and Case Studies and facilitate access for online analysis and downloading by industrial stakeholders.	12/31/22 Complete
FY23 Q2 (PNNL) Data Stakeholder Focus Groups	Provide Data Hub user accounts to at least 20 non-DOE stakeholders and conduct 3 focus group sessions to collect in-person feedback on the web portal's performance and utility.	3/31/23
FY23 Q3 (NREL/INL) Task 8 Case Study Support	LabKey "Study" home folders created for 4 new Task 8 Case Studies and two high-profile FCIC publications with data, metadata, and results integrated within the FCIC QbD database.	6/30/23
FY24 Q4 (PNNL) FCIC Data Stakeholder Workshop Planning	Agenda, guest speakers, and discussion sessions outlined for FCIC Data Stakeholder Workshop; Demonstrate QbD technology pathway critical properties rendering tool.	9/30/23
FY24 Q4 (PNNL) 3-year Final Deliverable	An online community of industry of stakeholders, National Lab researchers, academic partners, and BETO staff sharing data, knowledge, and tools is established on the FCIC Data Hub with a target of having 100+ active users from commercial entities.	9/30/24



## Responses to Previous Reviewers' Comments



- Comment, Reviewer 1: A huge undertaking within BETO that is clearly advancing towards its goal of a comprehensive data hub. From other presentations it appears that some datasets are going to be added onto this with different UI / layouts. If true, this would harm the overall effort of a consistent platform. The data needs to be accessed, entered, and viewed in a consistent manner to give people confidence that the data is accurate as well as to provide a consistency in quality required for such an effort....Recommend that only a small group of people be allowed to make data entries instead of anyone. Data can and will likely be entered sloppily by users in a rush to get the menial task over with. However, if a group of people had sole control, they could ensure that the data was equivalent, definitions were consistent, and experiments were taking the correct measurements. The project is not only for the final depository of data but could be used as a project planning tool as well. It has the potential to help guide projects towards the correct items to measure in their labs, and in their processes....This tool should be used as a specification for data collection for government projects. The same data, collected in the same ways, done on each project is the most important way of making this dataset valuable. A publicly available tool which can provide data on a particular feedstock at a particular point in a system that has been vetted by national lab scientists is an exciting advancement.
- Response: We agree that maximizing the value of the Bioenergy Data Hub will require researchers working across the nine National Labs of the FCIC to agree upon and adhere to common standards for material and process data and metadata to ensure that their uploaded datasets will be findable, accessible, interoperable, and reusable by our industry stakeholders. A key, measurable objective for our Task will be to harmonize datasets from FCIC technical reports and journal articles, and then upload them to the Data Hub to enable integrated views of FCIC data within a single interface with standardized metadata and units of measure. Workflows for achieving such data integration will necessarily involve only a small, select group of data managers from each of the FCIC National Labs to ensure data quality.



## Responses to Previous Reviewers' Comments



- Comment: The project is very ambitious in the amount of data and information it is encompassing. Standardizing QbD nomenclature as well as linking to other programs such as the Bioenergy Feedstock Library will be key to institutionalizing the concept across FCIC. Although this effort is great for FCIC internal collaboration, the challenge will be to get industry and even other technology areas of BETO to adopt the standardized nomenclature and start using the data sets and reports. The beta testing using the IAB and getting input from the DFO partners is a good start. Ease of use will be key to the adoption of this technology. The data set on High-Temperature Conversion does not have any DFO or industry partner using the pyrolysis technology for the production of biofuels. This will limit the applicability of the current data sets in this space. More extensive input from industry as well as Beta-testing by industry users will be key to determining the value to the biorefinery community.
- Response: Our uniform applications of Quality by Design (QbD) nomenclature within the Process Stream Critical Properties, Unit Operation Critical Process Parameters, and Technology Pathway Configuration data tables on the Data Hub are indeed standardizing the way we think and talk about biorefinery materials and processes within the FCIC. Further integration of the FCIC's QbD framework for biorefineries with the database schema of the BFL can form the basis of a new "Bioeconomy Feedstock Conversion Ontology" that provides a shareable and reusable framework of formally specified objects, classes, attributes, and relations to support biorefinery development and the creation of markets for renewable feedstocks, intermediates, and products.



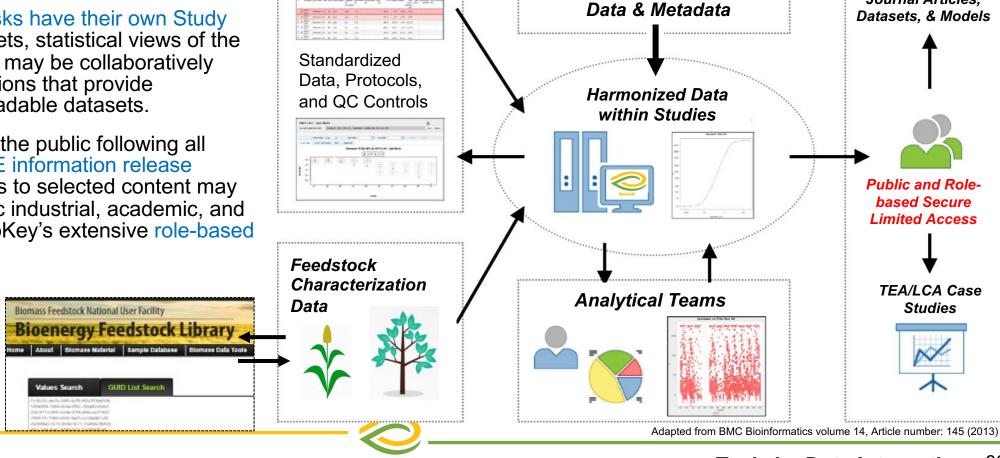


Journal Articles.

#### **LabKey Data Server Infrastructure:**

- We are leveraging LabKey's built-in workflows to integrate bench-to-PDU scale feedstock and experimental data within "Study" data structures.
- The FCIC's 34 Subtasks have their own Study folders wherein datasets, statistical views of the data, figures, and text may be collaboratively compiled into publications that provide standardized, downloadable datasets.
- Access is provided to the public following all National Lab and DOE information release polices; secure access to selected content may be provided to specific industrial, academic, and NL partners using LabKey's extensive role-based permissions\* system.

\*LabKey user roles include FCIC Data Hub Admins, Editors, Analysts (can create and enable script code) and Readers. Roles may be assigned on a per-folder basis for granular access permission assignment.



Experimental Teams

Experimental

(Assay) Data

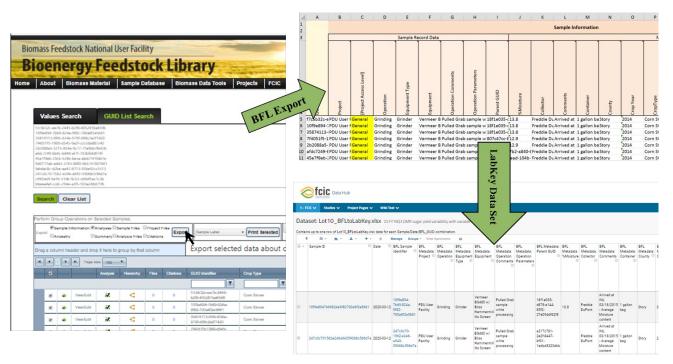
Management

## 4 – Progress: Bioenergy Feedstock Library integration via Python API





**Creation of LabKey Datasets directly from Bioenergy Feedstock Library (BFL) exports** 



#### **Description**

LabKey's Python API was used to develop an automated process for creating Study datasets for batches of sample provenance and analytical data exported from the Bioenergy Feedstock Library (BFL).

#### Value of new tool

The tool reduces the need to manually create redundant data structures between the two systems and represents an additional method to establish interoperability between the BFL and LabKey.

#### **Potential Customers & Outreach Plan**

INL researchers can more easily import samples and data directly into LabKey that are being tracked for the FCIC in the BFL database system.

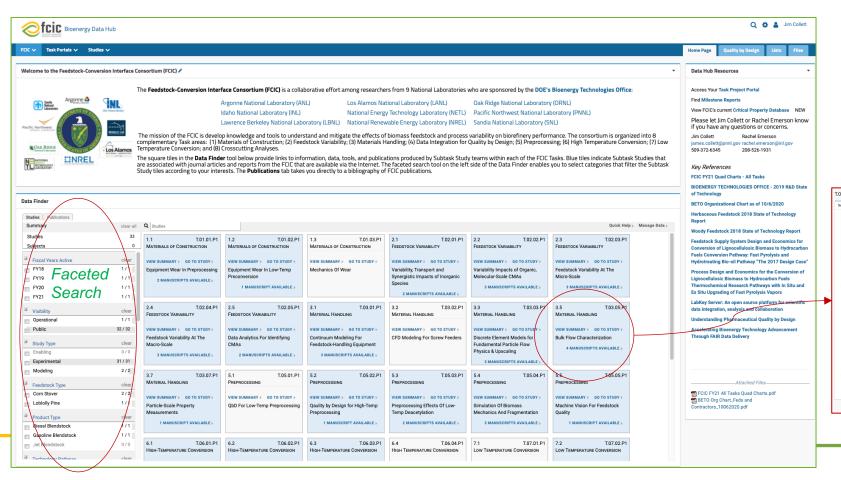


## 4 – Progress: Finding Knowledge and Tools on the FCIC Data Hub

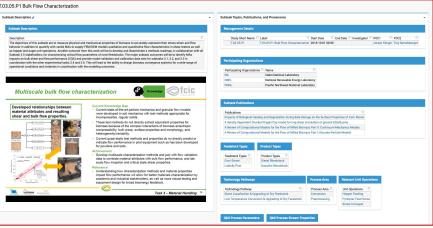




- The <u>Data Hub Web Portal</u> is now online and equipped with a <u>"Data Finder" dashboard</u> for accessing "Study" folders for each FCIC Subtask. FCIC publications may be accessed via the "Publications" tab.
- Data accruing on the Data Hub are easily Findable and Accessible (per FAIR data guidelines) via a <u>faceted search tool</u> that allows
  users to click on metadata tags to rapidly down-select to Subtasks covering specific R&D focus areas.

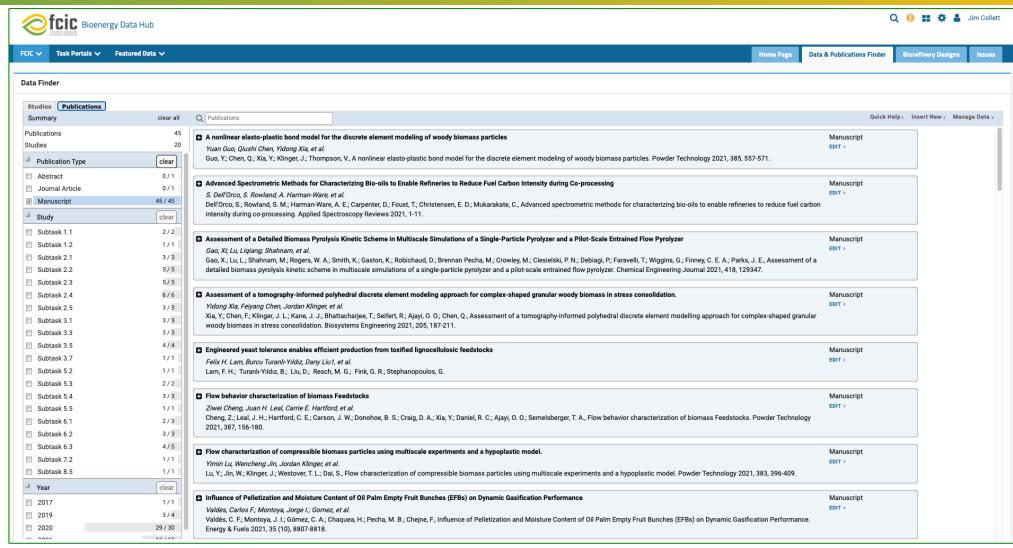


<u>Subtask metadata tags</u> now include a range of 260 descriptors for analytical methods, feedstock types, product types, technology pathways, process areas, unit operations, process parameters, process stream properties, fiscal years active, and participating organizations.



## Data Finder Publications Tab





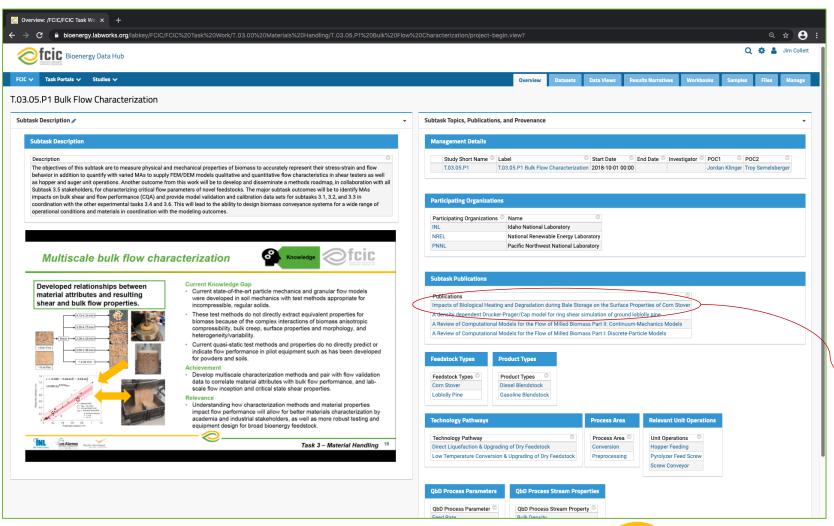
Detailed information about FCIC publications may be found under the Publications tab of the Data Finder.



### 4 – Progress: LabKey Subtask Studies

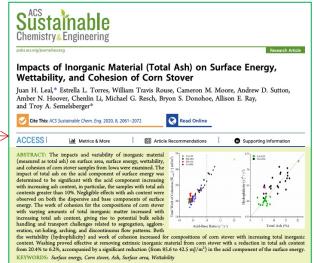






#### **Technical Approach:**

- Each of the FCIC's 35 Subtasks has its own "Study" home page on the Data Hub
- Studies have links for datasets, analyzed data views, results narratives, provenance, publications, and metadata associated with the Subtask.
- Subtask Studies provide supporting data and information for the 66 "Knowledge" and "Tool" products developed thus far by the FCIC.

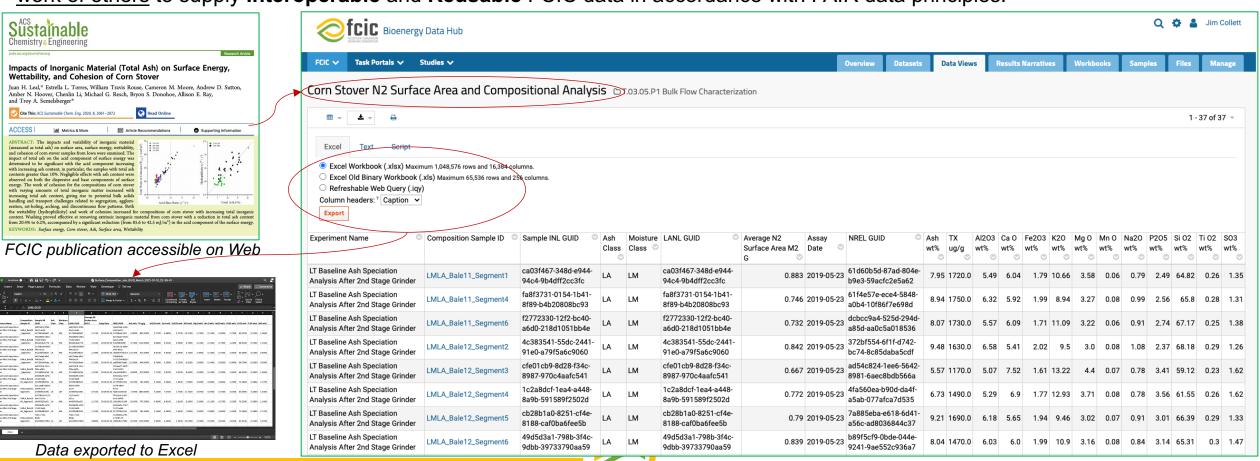


# 4 – Progress: Subtask Data Access and Export





- Harmonized, integrated datasets supporting FCIC publications may be <u>easily found</u>, filtered, sorted, and downloaded to Excel or to TSV or CSV text files.
- The integrated LANL, NREL and INL dataset shown below highlights how the Data Hub enables <u>self-service access to the past</u> work of others to supply **Interoperable** and **Reusable** FCIC data in accordance with FAIR data principles.

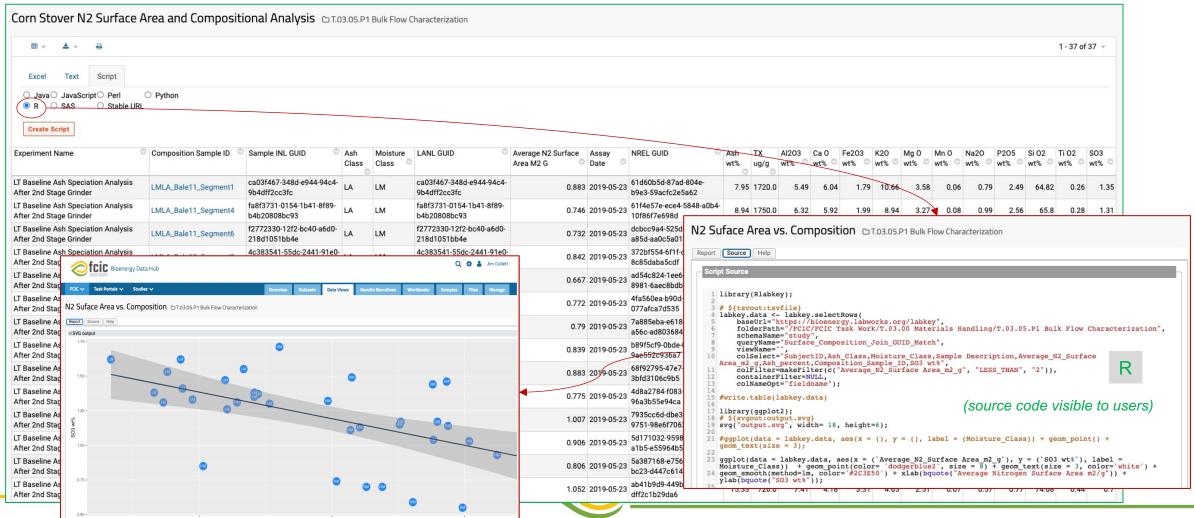


### 4 – Progress: Data Hub Analytical Tools





**Data Analysis:** The Task 4 Team is developing reusable data analysis and automated data processing tools using LabKey's extensive Application Programming Interfaces (APIs) for JMP/SAS, R, Python, Perl, Java, and JavaScript.



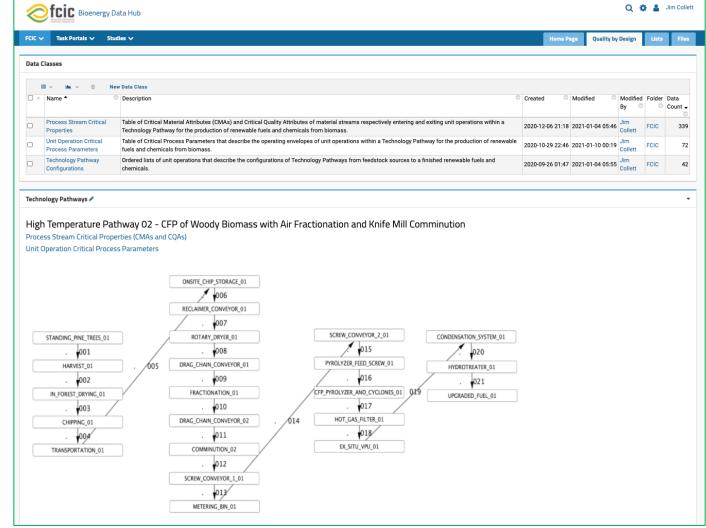
# 4 – Progress: Quality by Design Data Integration





## Mapping QbD Properties onto Technology Pathways

- Process Stream Critical Properties
- Unit Operation Critical Process Parameters
- Collaborative QbD property assignment supports:
  - self-organization of metadata into <u>ontologies</u> within the FCIC community.
  - consensus for defining <u>controlled vocabularies and</u> <u>units of measure</u> for use across the growing bioeconomy.





# 4 – Progress: Quality by Design Data Integration





- A <u>Process Stream Critical Properties</u> data table was created for managing the assignment of QbD critical material properties to feedstocks, intermediates, and products within the High- and Low-Temperature Conversion pathways.
- A related <u>Unit Operation Critical Process Parameters</u> data table was created for assignment of critical process parameters to unit
  operations within the High- and Low-Temperature Conversion technology pathways.
- Subtask teams and the FCIC PI have reached consensus on a first batch of 164 input material attributes, 85 output quality attributes, and 79 process parameters; these data are now being harmonized and uploaded to the database.

		<u> </u>						<u> </u>				
Ⅲ ▼ ▲ ▼ + ▼ 前 ★ ▼ ⊕ 1-12 of 12 ▼												
⊞ defa	## default This grid view has been modified. Revert Edit Save											
X ▼ Material Property IS ONE OF (MOISTU X ▼ Unit Op IS ONE OF (DRAG_CHAIN_CON Clear All												
0 =	Tech Pathway	Unit Op ▼	Intermediate Stream •	Intermediate State	Material Property ▼	-	CMA Upper	CMA Lower	CQA Upper	CQA Lower	Downstream Unit Op	Critical Downstream Unit Op
			0							Limit ©		
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	009	Chips	ASH_CONTENT_01	%	30.0	4.0	30.0	4.0	FRACTIONATION_01	COMMINUTION_01
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	009	Chips	PARTICLE_SIZE_DISTRIBUTION_01	mm	50.0	6.0	50.0	6.0	FRACTIONATION_01	COMMINUTION_01
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	009	Chips	MOISTURE_CONTENT_01	%	35.0	10.0	35.0	10.0	FRACTIONATION_01	COMMINUTION_01
□ 🖋 🙃	HIGH_TEMPERATURE_CONVERSION_02	FRACTIONATION_01	010	Chips	PARTICLE_SIZE_DISTRIBUTION_01	mm					DRAG_CHAIN_CONVEYOR_01	
	HIGH_TEMPERATURE_CONVERSION_02	FRACTIONATION_01	010	Chips	ASH_CONTENT_01	%					DRAG_CHAIN_CONVEYOR_01	
	HIGH_TEMPERATURE_CONVERSION_02	FRACTIONATION_01	010	Chips	MOISTURE_CONTENT_01	%					DRAG_CHAIN_CONVEYOR_01	
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	011	Chips	PARTICLE_SIZE_DISTRIBUTION_01	mm	50.0	6.0		1.0	COMMINUTION_02	PYROLYZER_FEED_SCREW_01
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	011	Chips	ASH_CONTENT_01	%	30.0	4.0		0.0	COMMINUTION_02	PYROLYZER_FEED_SCREW_01
	HIGH_TEMPERATURE_CONVERSION_02	DRAG_CHAIN_CONVEYOR_01	011	Chips	MOISTURE_CONTENT_01	%	35.0	10.0		3.0	COMMINUTION_02	PYROLYZER_FEED_SCREW_01
	HIGH_TEMPERATURE_CONVERSION_02	COMMINUTION_02	012	Particles	PARTICLE_SIZE_DISTRIBUTION_01	mm					SCREW_CONVEYOR_1_01	
	HIGH_TEMPERATURE_CONVERSION_02	COMMINUTION_02	012	Particles	ASH_CONTENT_01	%					SCREW_CONVEYOR_1_01	
	HIGH_TEMPERATURE_CONVERSION_02	COMMINUTION_02	012	Particles	MOISTURE_CONTENT_01	%					SCREW_CONVEYOR_1_01	
											·	



Data shown are not actual and are provided for layout demonstration only

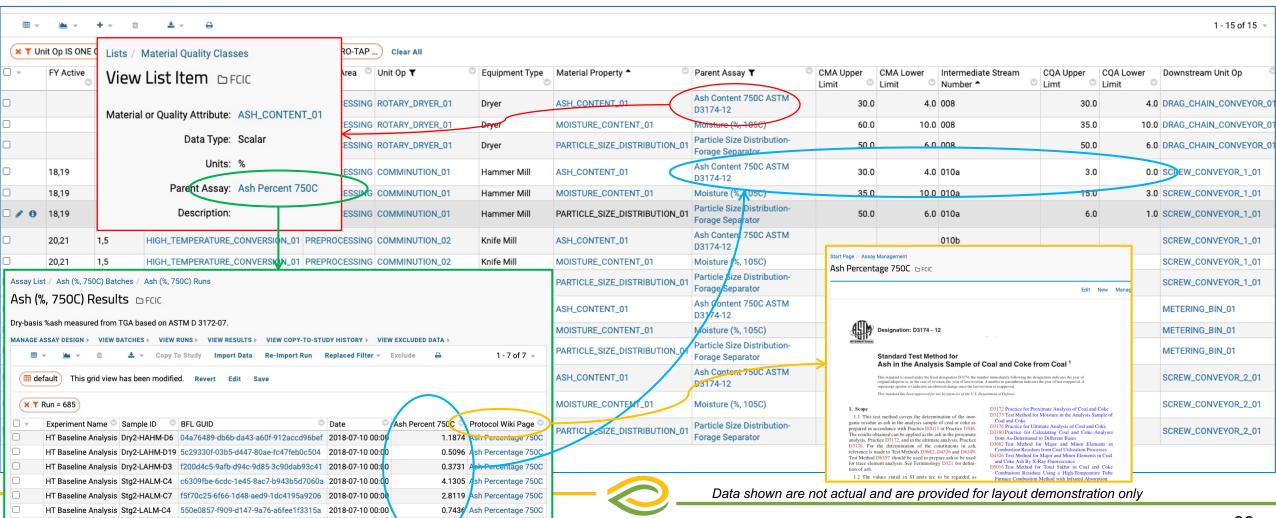
# 4 – Progress: Access to Data for QbD Critical Property Assignment

HT Baseline Analysis Stg2-LALM-C6 7599e95e-a198-8b48-af95-deb1b4021315 2018-07-10 00:00





• The QbD tables support links to experimental data that may be considered in the assignment of sets of critical properties to process intermediate streams or unit operations.



0.6884 Ash Percentage 750C

# 4 – Progress: Experimental Data Integration





- Screenshot shows part of a dataset containing 216 records with 224 data columns integrated via a simple SQL query from 3 uploaded data files containing:
  - Feedstock property data from INL
  - Bio-oil property data and pyrolyzer run metadata from NREL
  - Fuel property data and hydrotreater run metadata from PNNL
- Stable, reusable SQL code and user access to source files promotes trust in large datasets on the Data Hub for data mining analysis and machine learning.
- Performing such integrations manually in Excel is time-consuming and error-prone.

